Identification of Associations between Clinical Signs and Hosts to Monitor the Web for Detection of Animal Disease Outbreaks

Elena Arsevska, Unit for Control of Exotic and Emerging Diseases in Animals (UMR CMAEE), French Agricultural Research and International Cooperation Organization (CIRAD), Montpellier, France
Mathieu Roche, Unit for Land, Environment, Remote Sensing and Spatial Information (UMR TETIS), French Agricultural Research and International Cooperation Organization (CIRAD), Montpellier, France
Pascal Hendrikx, Unit for Coordination and Support to Surveillance (UCAS), French Agency for Food, Environmental and Occupational Safety (ANSES), Maisons-Alfort, France
David Chavernac, Unit for Control of Exotic and Emerging Diseases in Animals (UMR CMAEE), French Agricultural Research and International Cooperation Organization (CIRAD), Montpellier, France
Sylvain Falala, Unit for Control of Exotic and Emerging Diseases in Animals (UMR CMAEE), French National Institute for Agricultural Research (INRA), Montpellier, France
Renaud Lancelot, Unit for Control of Exotic and Emerging Diseases in Animals (UMR CMAEE), French Agricultural Research and International Cooperation Organization (CIRAD), Montpellier, France
Barbara Dufour, Ecole vétérinaire d’Alfort, EpiMAI USC Anses, Université Paris Est, Maisons-Alfort, France

ABSTRACT

In a context of intensification of international trade and travels, the transboundary spread of emerging human or animal pathogens represents a growing concern. One of the missions of the national veterinary services is to implement international epidemiological intelligence for a timely and accurate detection of emerging animal infectious diseases (EAID) worldwide, and take early actions to prevent their introduction on the national territory. For this purpose, an efficient use of the information published on the web is essential. The authors present a comprehensive method for identification of relevant associations between terms describing clinical signs and hosts to build queries to monitor the web for early detection of EAID. Using text and web mining approaches, they present statistical measures for automatic selection of relevant associations between terms. In addition, expert elicitation is used to highlight the most relevant terms and associations among those automatically selected. The authors assessed the performance of the combination of the automatic approach and expert elicitation to monitor the web for a list of selected animal pathogens.

KEYWORDS

Disease Emergence, Expert Elicitation, Information Retrieval, Ranking Function, Text Mining, Web Mining

1. INTRODUCTION

In recent years, the increased globalisation, movement of passengers and international trade has influenced the (re)emergence of new and exotic infectious diseases (Morens, Taubenberger, & Fauci, 2013). The traditional disease surveillance systems, organized via a multilevel health infrastructure,
show delays in reporting disease outbreaks, starting from the first observation of clinical signs, laboratory confirmation until public communication (Chan et al., 2010). In consequence, the delays in reporting of disease outbreaks have themselves delayed the implementation of disease control measures, and thus influenced the spread of pathogens to uninfected territories (Khomenko et al., 2013).

As a complement to the traditional disease surveillance systems, several event-based surveillance systems (web monitoring systems) gather information about infectious disease outbreaks from automatically retrieved articles from the web (Collier et al., 2010; Freifeld, Mandl, Reis, & Brownstein, 2008; Steinberger, Fuart, Best, Von Etter & Yangarber, 2008). For this purpose, the current web monitoring systems use a specific vocabulary, such as names of diseases and clinical signs. However, it is not clear how these systems identify the vocabulary to mine the web, and especially for animal infectious diseases.

Innovative, data mining approaches have been successfully applied to clinical records in human medicine (Chapman, Dowling, & Wagner, 2004; Friedlin, Grannis, & Overhage, 2008) and to articles retrieved from the web, such as with the web monitoring systems HealthMap and BioCaster (Brownstein, Freifeld, Reis, & Mandl, 2008; Collier et al., 2008). However, the data mining approaches in animal health face challenges such as multiple vertebrate and possibly invertebrate (insects, ticks…) hosts and less formal vocabulary (Santamaria & Zimmerman, 2011; Smith-Akin, Bearden, Pittenger, & Bernstam, 2007). Indeed, a single pathogen agent can affect multiple animal hosts at the same time (cattle, sheep, goats, pigs) and can manifest with similar or different clinical signs. Furthermore, the clinical signs can vary from very typical, specific clinical signs (congenital malformations and deformations, blister-like sores on the skin and mucous membranes, haemorrhagic syndrome etc.) to less typical, non-specific clinical signs (fever, weakness, diarrhoea, etc.).

In this paper, we propose and evaluate a new method that combines text and web mining approaches to select relevant associations between terms describing hosts and clinical signs to build queries to mine the web and detect an emergence of an infectious animal disease outbreak. We focus on the new and exotic infectious animal diseases.

After the presentation of the related work in the next section, sections 3 and 4 present the framework of our method and the text and web mining approaches for identification of associations between terms describing hosts and clinical signs. Sections 5 and 6 consider the experiments and the results of the evaluations. Finally, in the last section we discuss this paper and present our future work.

2. RELATED WORK

The Argus system, hosted at Georgetown University Medical Centre (USA), uses a simple method to detect articles on the web, using search terms of multilingual disease names (Nelson et al., 2010; Nelson et al., 2012).

More complete, web search criteria are proposed by the HealthMap team, which include disease names (scientific and common), clinical signs, keywords, and phrases. The terms originate from a dictionary of pathogens (human, plant, and animal diseases) and geographic names (country, province, state, and city). HealthMap integrates outbreak data from multiple electronic sources, including news feed aggregators (e.g., Google News), expert curated accounts (e.g., ProMED-mail), multinational surveillance reports (e.g., Eurosurveillance), and validated official alerts e.g., from WHO and OIE (Brownstein, Freifeld, Reis, & Mandl, 2008).

To detect articles on the web (e.g., Google News), the International Biosurveillance System (IBIS) uses the knowledge of registered users of the system that propose the search terms themselves. Registered users can edit the search terms and add or edit the tags in the articles that are relevant to the search terms (Lyon et al., 2013).

The GPHIN system developed by the Public Health Agency of Canada, retrieves automatically articles from news feed aggregators (e.g., Al Bawaba and Factiva), based on search terms and
Supervised Machine Learning for Plants Identification Based on Images of Their Leaves
www.igi-global.com/article/supervised-machine-learning-for-plants-identification-based-on-images-of-their-leaves/168499?camid=4v1a

Tools for Corporate Assessment of Sustainable Development
www.igi-global.com/chapter/tools-corporate-assessment-sustainable-development/51775?camid=4v1a